

# AMiner

**Jiawei Han**  
Abel Bliss Professor of Computer Science, University of Illinois  
data mining, database systems, data warehousing, information networks  
Verified email at cs.uiuc.edu - Homepage

Title	1-20	Cited by	Year
Data mining: concepts and techniques		29044	2011
J Han, M Kamber, J Pei			
Elsevier			
Mining frequent patterns without candidate generation	6391		2000
J Pei, J Han, Y Yin			
ACM Sigmod Record 29 (2), 1-12			
Data mining: an overview from a database perspective	2694		1996
MS Chen, J Han, P Yu			
Knowledge and Data Engineering, IEEE Transactions on 8 (6), 866-883			
E cient and E ective Clustering Methods for Spatial Data Mining	2499		1994
RT Ng, J Han			
Proc. of, 144-155			
PrefxSpan: Mining sequential patterns efficiently by prefix-projected pattern growth	2106		2001
J Pei, J Han, B Mortazavi-Asl, H Pinto, Q Chen, U Dayal, MC Hsu			
icdm, 0215			
Mining frequent patterns without candidate generation: A frequent-pattern tree approach	1851		2004
J Han, J Pei, Y Yin, R Mao			
Data mining and knowledge discovery 8 (1), 53-87			

**Google Scholar**

**Citation Indices**

Years	All	Since 2011
2008	113507	56901
2009	1136	96
2010	568	485
2011		

**Co-authors** View all... Xifeng Yan, Philip S. Yu, Xiaofei He, Deng Cai, Jianyong Wang, Yizhou Sun, Jing Gao, Laks V.S. Lakshmanan, Guozhu Dong, Charu Aggarwal, Ke Wang, ChengXiang Zhai, Anthony KH Tung

**Papers** 790

**Lectures** 13

**Patents** 1

**AMiner** Whatever comes to your mind

**Jiawei Han (韓家炜)**

Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana-Champaign, Illinois Area | Computer Software

Professor (217) 333-6903 hanj@cs.uiuc.edu http://www.cs.uiuc.edu/~hanj/ External Links

Send a message https://www.linkedin.com/in/jiawei

Background Experience Professor UIUC August 2001 – Present (14 years 7 months)

Research Interests Data-Mining, Machine Learning, Information Extraction, Text Mining

1985 1990 1995 2000 2005 2010 2015

Activity #Papers #Citation H-Index G-Index Sociability

1 Systems and Methods for Detecting a Novel Data Class

Mohammad Mehedy Masud, Latifur Rahman Khan, Bhavani Marienne Thuraisingham, Qing Chen, Jing Gao, Jiawei Han

Publication-date: 2012-03-01 Application-date: 2011-08-22

Troubleshooting interactive complexity bugs in wireless sensor networks using data mining techniques

Mohammad Maifi Hasan Khan, Hieu Khac Le, Hossein Ahmadi, Tarek F. Abdelzaher, Jiawei Han

ACM Transactions on Sensor Networks (TOSN) (2014)

BibTeX http://dx.doi.org/10.1145/2530290

787

Bringing Structure to Text: Mining Phrases, Entity Concepts, Topics, and Hierarchies

Wei Shen, Jianyong Wang, Jiawei Han

Knowledge and Data Engineering, IEEE Transactions on (2014)

dx.doi.org/10.1109/TKDE.2013.3192

Entity Linking with a Knowledge Graph

Wei Shen, Jianyong Wang, Jiawei Han

Knowledge and Data Engineering, IEEE Transactions on (2014)

dx.doi.org/10.1109/TKDE.2013.3192

20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), New York 2014

dx.doi.org/10.1107/978-1-4419-6045-0\_12

14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), Las Vegas 2008

Watch + Watch +

# Discovering Expertise Semantics

- Quantifying researchers' expertise using publications

818  
Entity Linking with a Knowledge Base: Issues, Techniques, and Solutions  
Wei Shen, Jianyong Wang, **Jiawei Han**  
Knowledge and Data Engineering, IEEE Transactions (2015)  
[BibTex](#) [DOI](http://dx.doi.org/10.1109/TKDE.2014.2327028)

817  
Power-Based Diagnosis of Node Silence in Remote High-End Sensing Systems.  
Yong Yang, Lu Su, Mohammad Maifi Hasan Khan, Michael LeMay, Tarek F. Abdelzaher, **Jiawei Han**  
TOSN (2015)  
[BibTex](#) [DOI](http://doi.acm.org/10.1145/2661639)

816  
A Framework of Mining Trajectories from Untrustworthy Data in Cyber-Physical System.  
Lu An Tang, Xiao Yu, Quanquan Gu, **Jiawei Han**, Guofei Jiang, Alice Leung, Thomas F. La Porta  
TKDD (2015)  
[BibTex](#) [DOI](http://doi.acm.org/10.1145/2700394)

815  
A Unifying Framework of Mining Trajectory Patterns of Various Temporal Tightness.  
Jae-Gil Lee, **Jiawei Han**, Xiaolei Li  
IEEE Trans. Knowl. Data Eng. (2015)  
[BibTex](#) [DOI](http://dx.doi.org/10.1109/TKDE.2014.2377742)

814  
ePeriodicity: Mining Event Periodicity from Incomplete Observations  
Zhenhui Li, Jingjing Wang, **Jiawei Han**  
Knowledge and Data Engineering, IEEE Transactions (2015)  
[BibTex](#) [DOI](http://dx.doi.org/10.1109/TKDE.2014.2365801)





**Jiawei Han (韩家炜)** 

H-Index: 133 | #Paper: 818 | #Citation: 111164

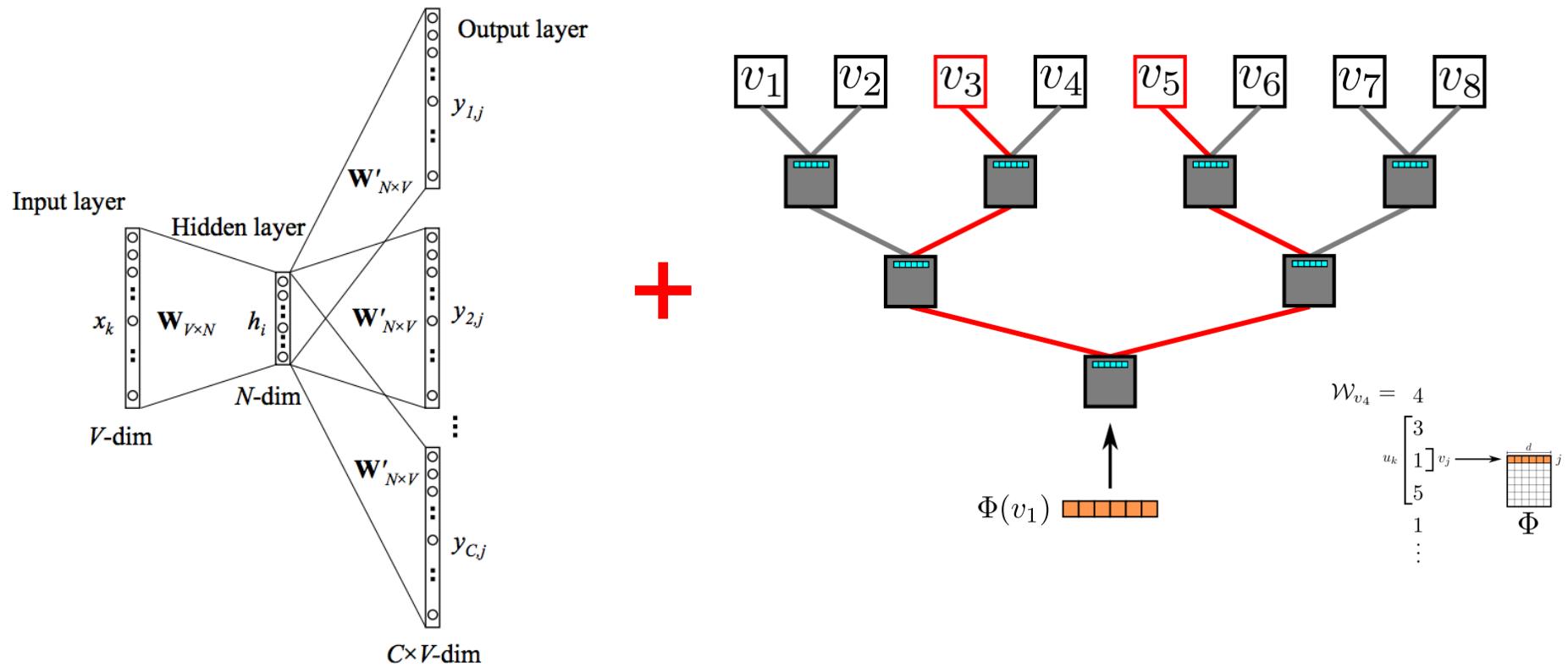
Department of Computer Science, University of Illinois at Urbana-Champaign  
Professor

Data Mining Information Extraction Data Analysis Machine Learning Text Mining

**Following | 87**

The straightforward method is to extract high frequent terms as expertise; however, the extracted terms may be not good terms to represent expertise.

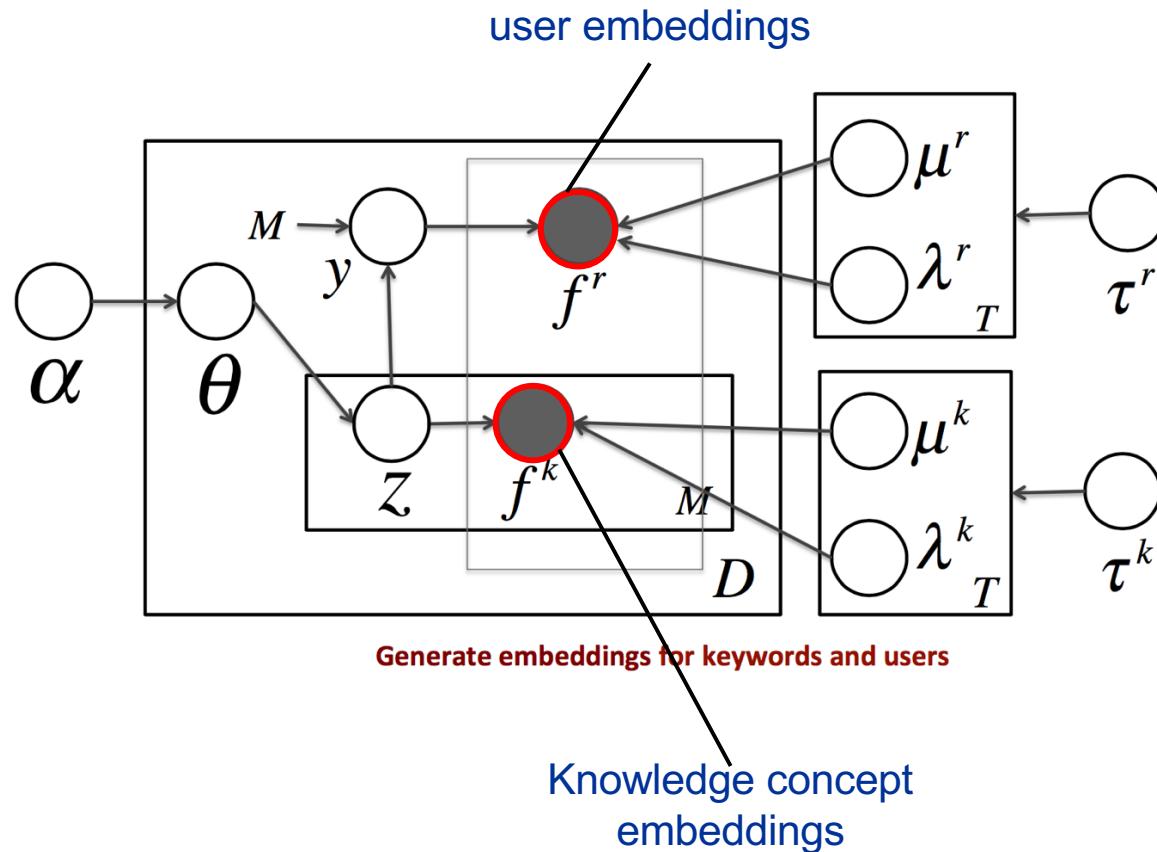
# Embedding Knowledge and Networks



Knowledge concept embedding

Network-based researcher embedding

# GenVector—bridging researcher network and knowledge graph



# Expertise-oriented Semantics in AMiner



Jiawei Han (韩家炜)

Following | 87

H-Index: 133 | #Paper: 818 | #Citation: 111164

📍 Department of Computer Science, University of Illinois at Urbana-Champaign

💼 Professor

Data Mining Information Extraction Data Analysis Machine Learning Text Mining

Similar

31809 views



Philip S. Yu

Following | 46

H-Index: 132 | #Paper: 857 | #Citation: 78606

📍 Department of Computer Science, University of Illinois Chicago

💼 Professor and Wexler Chair in Information Technology

Distributed System Query Optimization Query Processing Database Systems

Similar

1332 views



Jian Pei

Follow | 10

H-Index: 70 | #Paper: 297 | #Citation: 33670

📍 School of Computing Science, Simon Fraser University

💼 Professor

Information Extraction Data Mining Information Retrieval Visual Analytics

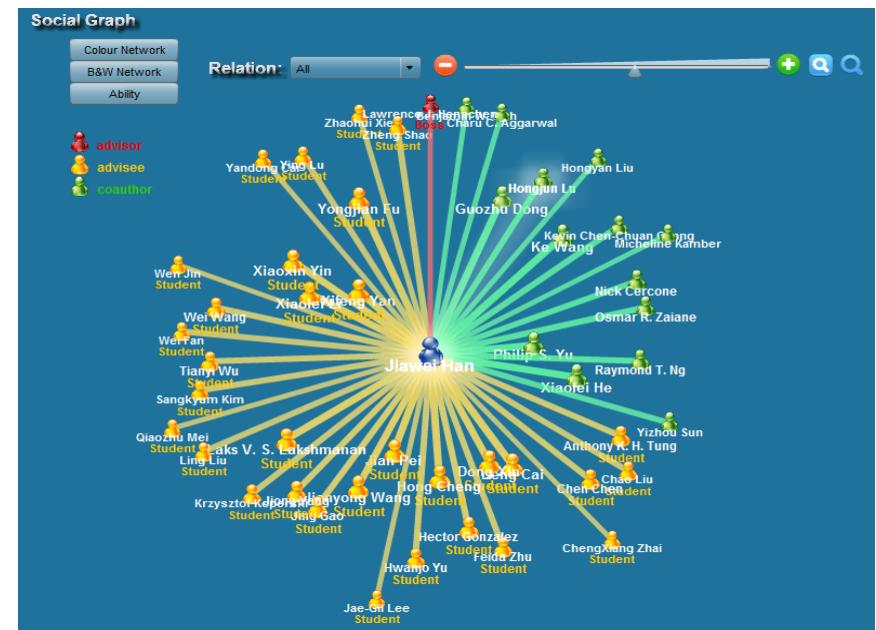
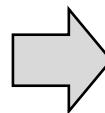
Similar

Document Clustering

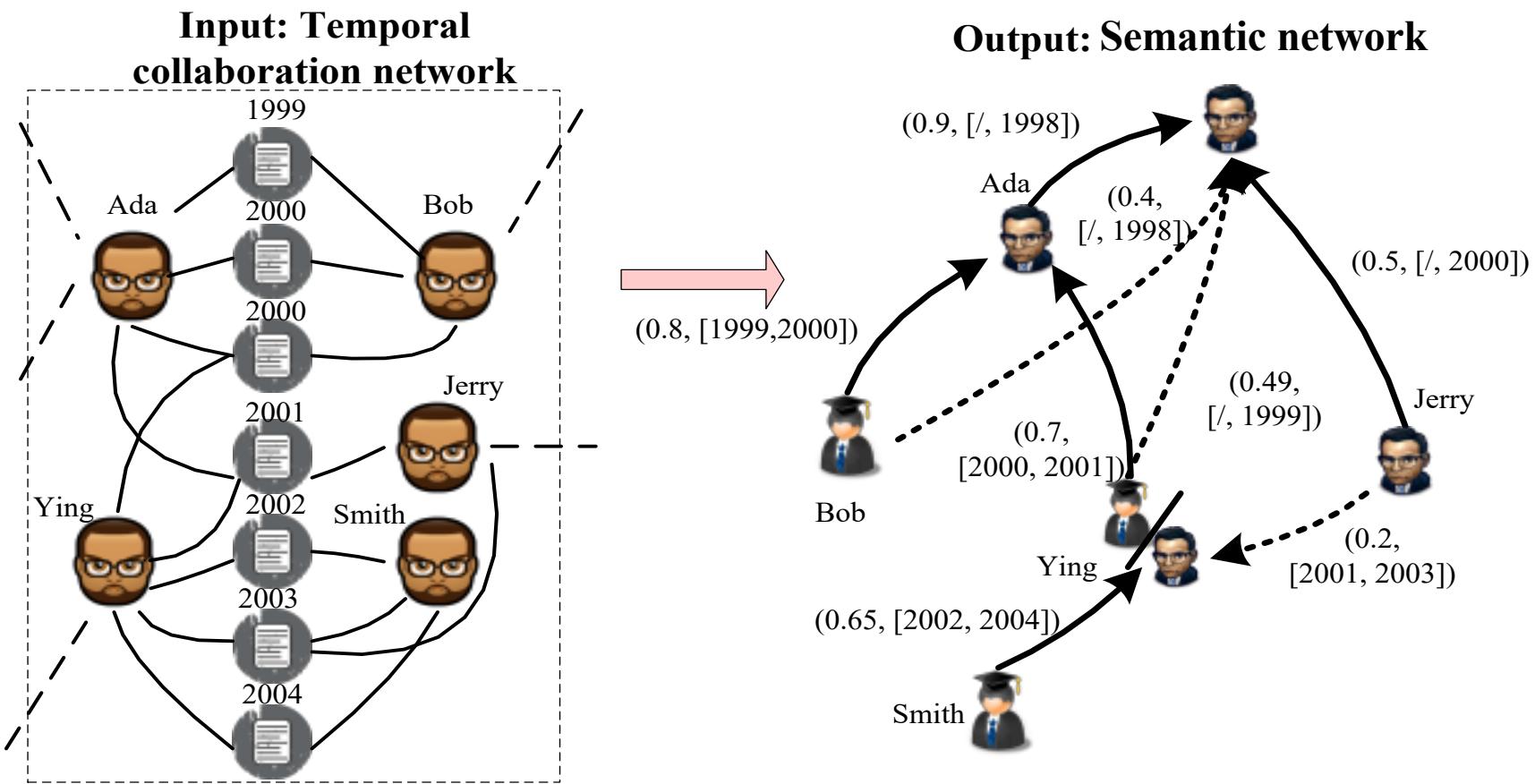
711 views

# Discovering Network Semantics

- How to mine semantics within networks?
  - Who are Jiawei Han's students and advisor?



# Mining Network Semantics



# AMiner

AMiner

Search

Jiawei Han (韓家炜)

Department of Computer Science, University of Illinois at Urbana-Champaign

Professor

(217) 333-6903

hanj@cs.uiuc.edu

<http://www.cs.uiuc.edu/~hanj/>

External Links

Research Interests

1985 1990 1995 2000 2005 2010 2015

About

- Papers 790
- Lectures 13
- Patents 1

All (790)	Recent 20	2015	2014	2013	2012	2011	2010	2009	2008	2007	2006	2005	2004	2003	2002	2001	2000	1999	1998	1997	1996	1995
1994	1993	1992	1991	1990	1989	1988	1987	1986	1985													

**Ego Network**

Similar Authors | Ego Network

**Entity Linking with a Knowledge Base: Issues, Techniques, and Solutions** Cited by 4

Wei Shen, Jianyong Wang, **Jiawei Han**

Knowledge and Data Engineering, IEEE Transactions (2015)

<http://dx.doi.org/10.1109/TKDE.2014.2327028>

**Mining Graph Patterns.** Cited by 20

Hong Cheng, Xifeng Yan, **Jiawei Han**

Frequent Pattern Mining (2014)

[http://dx.doi.org/10.1007/978-1-4419-6045-0\\_12](http://dx.doi.org/10.1007/978-1-4419-6045-0_12)

**Troubleshooting interactive complexity bugs in wireless sensor networks using data mining techniques** Cited by 5

Mohammad Maifi Hasan Khan, Hieu Khac Le, Hossein Ahmadi, Tarek F. Abdelzaher, **Jiawei Han**

ACM Transactions on Sensor Networks (TOSN) (2014)

<http://dx.doi.org/10.1145/2530290>

31



# 智能服务

# 人才大数据 – NSFC专家系统



EN | Dashboard | Login

## 国家自然科学基金委员会专家PROFILE系统 (试用版)

Whatever comes to your mind

国家自然科学基金委员会  
专家库

杰出青年基金获得者

包含中国国家自然科学基金杰出青年基金获得者  
相关信息的数据库。

Yu Jing 2016-07-01

Yu Jing 2016-02-18

Search Results for experts (2568 results)

李亚栋 (Yadong Li)  
H-Index: 42 | #Paper: 289 | #Citation: 6237  
清华大学

Nanostructures Hydrothermal Synthesis Nanocrystals Crystal Structure Semiconductors Self-Assembly Luminescence Nanoparticles

魏飞 (Fei Wei)  
H-Index: 40 | #Paper: 384 | #Citation: 5806  
清华大学

Carbon Nanotubes Chemical Vapor Deposition Carbon Nanotube Fluidized Bed Propylene Riser A. Carbon Nanotubes Sapo-34

薛其坤 (Qikun Xue)  
H-Index: 33 | #Paper: 281 | #Citation: 5792  
清华大学

Scanning Tunneling Microscopy Molecular Beam Epitaxy Topological Insulator Scanning Tunneling Microscope Silicon Angle-Resolved Photoemission Spectroscopy Electronic Structure Superconductivity

王建勇 (Jianyong Wang)  
H-Index: 27 | #Paper: 100 | #Citation: 5512  
清华大学

Efficient Algorithm XML Document Keyword Search Efficient Mining Keyword Query Web Data XML Query Extensive Experimental Study

李景虹 (Jinghong Li)  
H-Index: 38 | #Paper: 166 | #Citation: 5139  
清华大学

Graphene Electrocatalysis Biosensor Direct Electrochemistry Electrochemistry Dna Horseradish Peroxidase Ionic Liquid

唐杰 (Jie Tang)  
H-Index: 33 | #Paper: 185 | #Citation: 4596  
清华大学

Social Network Semantic Web Social Influence Heterogeneous Network Ontology Mapping Name Disambiguation Topic Model Unified Approach

NSFC学科代码

导出

NSFC

NSFC

NSFC

NSFC

NSFC

NSFC

NSFC

NSFC

NSFC

# 人才大数据 – 科技部人才评价

为科技部在库专家构建专家画像库

专家评价信息

姓名: 性别: 男 出生日期:  
单位: 职务: 职称: 教授  
研究方向: 强场物理 高能量密度物理 超快电子衍射与成像 实验室天体物理 激光物理

科技部人才中心

**总体评价**  
影响力10分  
根据荣誉奖励、主要任职、人才计划进行衡量，取单项最高分  
活跃度10分  
根据承担项目、发表论文、授权专利进行衡量，取单项最高分

**荣誉奖励**

**主要任职**  
中国科学院物理研究所研究员;

**人才计划**  
国家杰出青年科学基金

**承担项目**

构建专家画像库

# 人才大数据 – 阿里巴巴人才地图

阿里巴巴 学术资源地图 Data Mining 搜索 登录

搜索范围: ALL IEEE Fellow(2013-2016) ACM Fellow 英国皇家科学院 – Research Fellows Directory 英国皇家科学院 – Fellows Directory 中国科学院院士  
美国科学院外国专家

过滤条件: ALL h指数: >=60 (68) 50-59 (43) 40-49 (75) 30-39 (137) 20-29 (222) 10-19 (357) <10 (98)

语言: Chinese (259) English (196) Greek (39) French (29) German (28) Japanese (26) Indian (24) Korean (14) Italian (1)

国家: USA (229) China (111) United Kingdom (19) Germany (1) France (1) Australia (1) Canada (1) Italy (1) Spain (1) Netherlands (1) Sweden (1) Norway (1) Finland (1) Poland (1) Czech Republic (1) Hungary (1) Romania (1) Bulgaria (1) Turkey (1) Greece (1) Israel (1) Jordan (1) Lebanon (1) Saudi Arabia (1) Oman (1) Yemen (1) Iraq (1) Kuwait (1) Bahrain (1) Qatar (1) United Arab Emirates (1) Pakistan (1) India (1) Bangladesh (1) Sri Lanka (1) Maldives (1) Thailand (1) Philippines (1) Indonesia (1) Malaysia (1) Singapore (1) Vietnam (1) Mongolia (1) South Korea (1) Japan (1)

综合排序 IF H-index 学术活跃度

专家地图:

按照层级显示: 自动 大区 国家 城市 机构

专家地图显示了全球学术人才分布情况。图中以国家为单位，周围环绕着不同大小的同心圆，圆的大小代表该国家内专家的数量。例如，中国（中心点）有87位专家，日本有39位专家，印度有25位专家等。右侧详细展示了聂再清（Zaiqing Nie）的信息，他是中国的一位研究员，拥有19的H指数，工作于Microsoft Research Asia。

聂再清 (Zaiqing Nie) H-index: 19 Senior Researcher / Research Manager Microsoft Research Asia

Zaiqing Nie 聂再清

Jiawei Han (Jiawei Han) H: 152 A: 430.90 Abel Bliss Professor Department of Computer Science (217) 333-6903 hanj@illinois.edu

Philip S. Yu (Philip S. Yu) H: 144 A: 457.58 Distinguished Professor Department of Computer Science (312) 996-0498 psyu@uic.edu

Mohammed J. Zaki (Mohammed J. Zaki) H: 61 A: 22.81 Professor Department of Computer Science (518) 276-6340 zaki@cs.rpi.edu

Jian Pei (Jian Pei) H: 75 A: 61.60 Professor School of Computing 1-778-782-6851, 1-778 jpei@cs.edu.ca, jpei

# 人才大数据 – CCF专家系统



CCF 专家库



首页

学会活动

活动统计

设置

用户管理

角色管理

协办单位

贡献类别

活动类型

机构列表

人工智能

搜索

搜索范围: CCF会员 会士(F) 杰出会员(D) 高级会员(S) 全球专家

过滤条件: CCF会员

h指数: >=60(2) 50-59(2) 40-49(2) 30-39(13) 20-29(11) 10-19(46) <10(123)

语言: Chinese(187)

国家: China(164) Singapore(1)

相关度

学术成就 学术活跃度 领域新星 学会贡献

刘大有 (Dayou Liu)

0.00 23 19.57

吉林大学计算机科学系/计算机科学与技术学院

liudy@jlu.edu.cn

研究兴趣:

人工智能 定性空间推理 遗传算法 复杂网络 约束满足问题  
数据挖掘 计算机应用 时空推理

欧阳丹彤 (Dantong Ouyang)

0.00 10 1.73

吉林大学计算机科学与技术学院

ouyd@jlu.edu.cn

研究兴趣:

基于模型的诊断 人工智能 基于模型诊断 启发式搜索  
离散事件系统 混合键合图 约束满足问题 极小碰集

宗成庆 (Chengqing Zong)

0.00 18 37.97

Professor 中科院自动化研究所  
+86-10-8254 4688  
cqzong@nlpr.ia.ac.cn

研究兴趣:

机器翻译 人工智能 自然语言处理 条件随机场  
中文信息处理 机器翻译评测 计算机应用 命名实体

刘群 (Qun Liu)

0.00 31 55.10

研究兴趣:

# AMiner应用于搜狗搜索

建立常用搜索同义词  
库及上万节点的知识  
图谱

## 搜狗学术搜索提供者



搜狗学术

data mining, machine learning, social network, deep learning, healthcare

搜狗搜索

### 2016 Most Influential Scholars (Computer Science)



Timothy L. Harris

h-index: 40 | 论文数: 75 | 引用数: 16931

Lead Researcher

Oracle Labs group in Cambridge, UK

Transactional Memory Software Transactional Memory



Peter Druschel

h-index: 68 | 论文数: 196 | 引用数: 38498

Professor

Distributed Systems Group Max Planck Institute for Software...

Operating System Overlay Network P2p ...



Andrew Warfield

h-index: 24 | 论文数: 56 | 引用数: 14700

Associate Professor

Department of Computer Science University of British Columbia

Virtual Machine Virtualization High Availability ...



Michael Burrows

h-index: 38 | 论文数: 90 | 引用数: 20023

Computer Scientist

Google

Distributed System Access Control Rocky Shore



Keir Fraser

h-index: 16 | 论文数: 24 | 引用数: 14297

Co-Founder and Chief Architect

Coho Data

Operating System Virtual Machine Monitor ...

### Academic Rankings

#### Researcher Rank

Rank researchers by various metrics.

#### Organization Rank

Rank organizations by different metrics.

#### Conference Rank

Rank conferences in Computer Science by Impact Factor.

#### Best Papers & Top Cited Papers

Best papers and top cited papers in Computer Science.

### AI LABS

#### Gender Prediction

Predict gender with people name and affiliation/location.

#### Open Data

Citation network, topic experts, disambiguated names, etc.

#### Open Seminar

Recent academic seminars, e.g., all seminars from Tsinghua.

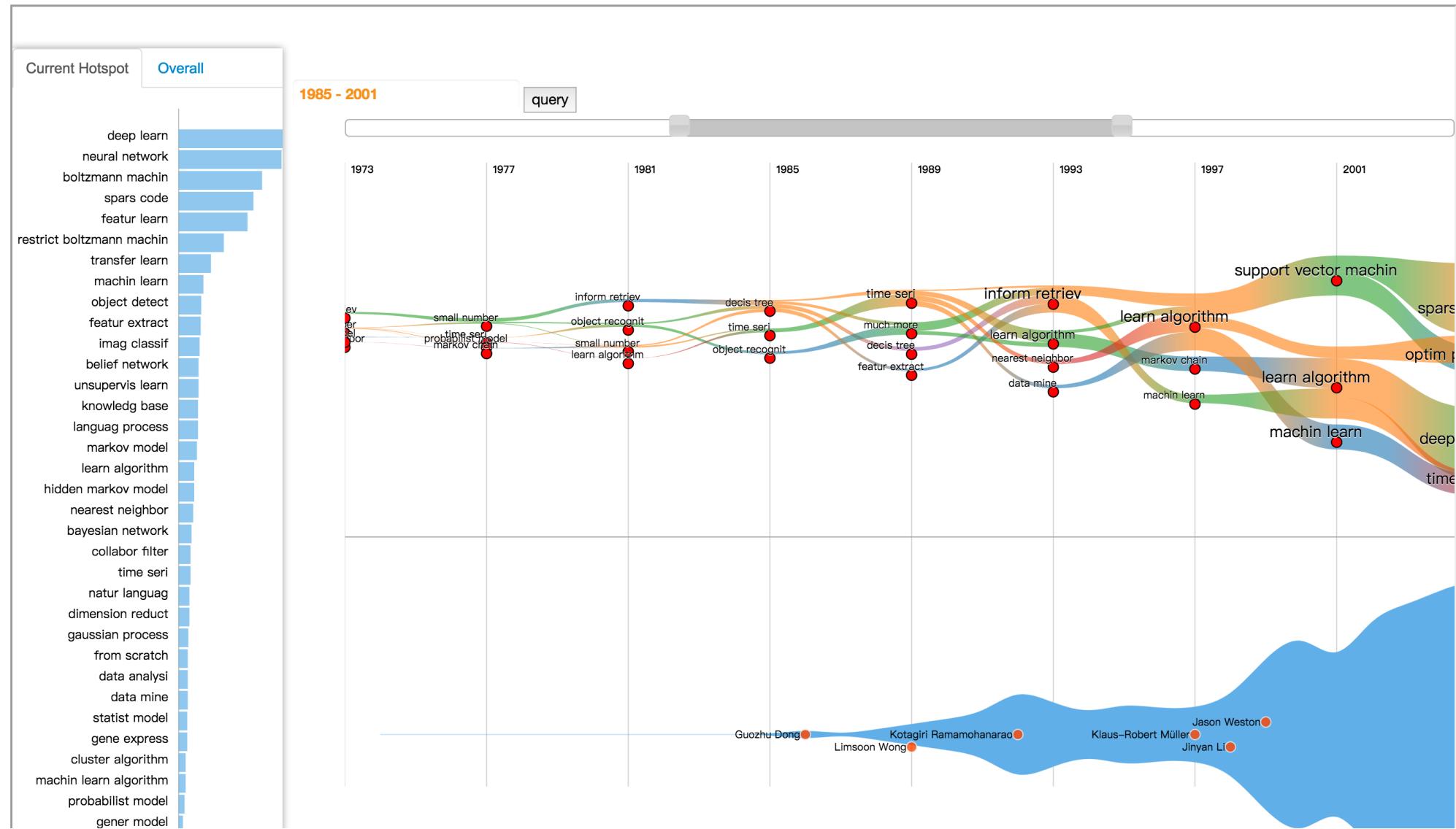
#### Open API

APIs to access our data and to build advanced functions.

# 全球顶级人才迁徙图

# 知识图谱 – 技术趋势预测

技术趋势预测:



# 知识图谱 – 技术趋势预测

## 知识图谱

- L0 Design and analysis of algorithms (算法的设计与分析) # 11
- L1 Data structures design and analysis (数据结构的设计与分析) # 5
- L2 Pattern matching (模式匹配) # 20
- L3 data mining**
  - L3 pattern matching
  - L3 lempel-ziv compressor
  - L3 geometry and discrete structures
  - L3 indexing data structure
- L0 Visualization (可视化) # 6
- L1 Visualization application domains (可视化应用程序域) # 4
- L2 Scientific visualization (科学可视化)**  
ALIAS: scientific visualisation,
  - L3 reporting system
  - L3 data mining**
    - L3 transformation and loading
    - L3 bioinformatics
    - L3 extraction
- L0 Probability and statistics (概率和统计) # 9

**INFO EXPERTS PUBLICATIONS MORE >**



**刘冰 (Bing Liu)**

H 73 C 41056 P 272

Professor  
Department of Computer Science, University of Illinois at Chicago (UIC)  
研究兴趣:

Data Mining, Opinion Mining, Web Pages, Web Mining, Association Rule, Proposed Technique, Knowledge Discovery, Web Page



**Wynne Hsu**

H 47 C 17021 P 196

Provost's Chair; Professor  
Department of Computer Science, School of Computing, National University of Singapore  
研究兴趣:

Data Mining, Indexation, Association Rule Mining, Association Rule, Data Structure, Experiment Result, Feature Extraction, Data Cleaning



**Tok Wang Ling**

H 35 C 4741 P 256

Professor  
Department of Computer Science School of Computing National University of Singapore  
研究兴趣:

Xml Document, Xml, Xml Database, Xml Data, Data Model, Entity Relationship, Keyword Search

# Open Data — Citation Network

<https://aminer.org/citation>

AMiner Whatever comes to your mind  EN  Dashboard |  

# Citation Network Dataset

Citation Network Dataset

Overview Description References

## Overview

### Overview

The data set is designed for research purpose only. The citation data is extracted from DBLP, ACM, and other sources. The first version contains 629,814 papers and 632,752 citations. Each paper is associated with abstract, authors, year, venue, and title.

The data set can be used for clustering with network and side information, studying influence in the citation network, finding the most influential papers, topic modeling analysis, etc.

A larger version will be released soon.

**Citation-network V1:** 629,814 papers and >632,752 citation relationships (2010-05-15).

**Citation-network V2:** 1,397,240 papers and >3,021,489 citation relationships (2010-09-13).

**DBLP-Citation-network V3:** 1,632,442 papers and >2,327,450 citation relationships (2010-10-22).

**DBLP-Citation-network V4:** 1,511,035 papers and 2,084,019 citation relationships (2011-01-08).

**DBLP-Citation-network V5:** [download from mirror site] 1,572,277 papers and 2,084,019 citation relationships (2011-01-08).

**DBLP-Citation-network V6:** 2,084,055 papers and 2,244,018 citation relationships (2013-09-29).

**DBLP-Citation-network V7:** [download from mirror site] 2,244,021 papers and 4,354,534 citation relationships (2014-05-25).

**ACM-Citation-network V8:** 2,381,688 papers and 10,476,564 citation relationships (2016-04-02)

Data set	#paper	#Citation Relationship	Comment
Citation-network V1	629,814	>632,752	
Citation-network V2	1,397,240	>3,021,489	
DBLP-Citation-network V3	1,632,442	>2,327,450	
DBLP-Citation-network V4	1,511,035	2,084,019	Arnetminer [2011-01-08]
DBLP-Citation-network V5 [download from mirror site]	1,572,277	2,084,019	Arnetminer [2011-02-21]
DBLP-Citation-network V6	2,084,055	2,244,018	Arnetminer [2013-09-29]
DBLP-Citation-network V7 [download from mirror site]	2,244,021	4,354,534	Arnetminer [2014-05-25]
ACM-Citation-network V8	2,381,688	10,476,564	Arnetminer [2016-04-02]

# Open Data – Network with Rich Semantics

Extraction and Mining of  
Academic Social Networks

Overview  
Data Description  
References

## Overview

This dataset is designed for research purpose only.

The content of this data includes paper information, paper citation, author information and author collaboration. **2,092,356** papers and **8,024,869** citations between them are saved in the file [AMiner-Paper.rar](#); **1,712,433** authors are saved in the file [AMiner-Author.zip](#) and **4,258,615** collaboration relationships are saved in the file [AMiner-Coauthor.zip](#).

FileName	Node	Number	Size
<a href="#">AMiner-Paper.rar</a> [download from mirror site]	Paper	2,092,356	509 MB
	Citation	8,024,869	
<a href="#">AMiner-Author.zip</a> [download from mirror site]	Author	1,712,433	167 MB
<a href="#">AMiner-Coauthor.zip</a> [download from mirror site]	Collaboration	4,258,615	31.5 MB

<https://aminer.org/AMinerNetwork>

:: The relationship between author id and paper id [AMiner](#)-

# Open Data – Open Academic Graph (OAG)

<https://aminer.org/open-academic-graph>

<https://www.openacademic.ai/>



ABOUT NEWS EVENTS MEMBERS PROJECTS DATA & TOOLS

## News

### Open Academic Graph is announced!

This data set is generated by linking two large academic graphs: **Microsoft Academic Graph (MAG)** and **AMiner**, and it is used for research purpose only. This version includes **166,192,182** papers from MAG and **154,771,162** papers from AMiner. We generated **64,639,608** linking (matching) relations between the two graphs. In the future, more linking results, like authors, will be published. It can be used as a unified large academic graph for studying citation network, paper content, and others, and can be also used to study integration of multiple academic graphs.

The overall data set includes three parts, which are described in the table below:

Data Set	Download Link	#Paper	Total Size	Date
Linking relations (matching)	<a href="#">Coming Soon</a>	64,639,608	1.6GB	2017-06-22
MAG papers	<a href="#">Coming Soon</a>	166,192,182	104GB	2017-06-09
AMiner papers	<a href="#">Coming Soon</a>	154,771,162	39GB	2017-03-22



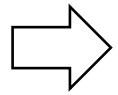
Edward Feigenbaum  
专家系统之父  
图灵奖



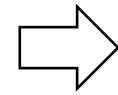
Tim Berners Lee  
WWW创始人  
图灵奖

# AMiner: 知识驱动的科技情报挖掘

大数据



知识



智能

\*人工智能的两个重要阶段：大规模知识库 + 智能服务

[1] J. Tang, J. Zhang, L. Yao, J. Li, L. Zhang, and Z. Su. ArnetMiner: Extraction and Mining of Academic Social Networks. KDD'08. pp.990-998.



# AMiner: 知识驱动的科技情报挖掘

<http://aminer.org>

## Thanks to our partners



# AMiner: 知识驱动的科技情报挖掘

<http://aminer.org>



[1] J. Tang, J. Zhang, L. Yao, J. Li, L. Zhang, and Z. Su. ArnetMiner: Extraction and Mining of Academic Social Networks. KDD'08. pp.990-998.